Session title: Can better use be made of public data for example in health research?

Onora O’Neill

Is Current Data Protection Legislation Coherent?

Data protection legislation assumes that we can protect privacy by regulating all uses of specific types of information (content), and more specifically that it is possible to partition information about persons exhaustively into the personal and the non personal. Other approaches to protecting privacy are based on the thought that we can best do so by regulating types of communication, and do not require us to partition information into the personal and non-personal. I shall contrast the two approaches and consider which might offer a more feasible and effective approach to protecting personal privacy.

I shall comment on ways in which data protection is *meant* to work, especially in biomedical contexts, and argue that it is unfit for purpose. This is a blunt comment, made by a non lawyer. I thought I might as well warn you!

1. A Selective Approach to Privacy Protection

The UK has signed up to the European Convention on Human Rights (ECHR), but has not implemented Art. 8, the Convention right to privacy,[1] in any systematic way. We do however have the *Data Protection Act 1998 (DPA)* which seeks to protect aspects of informational privacy. The Act is based on the European Data Protection Directive of 1995 [2] and aims to regulate the use of **personal information** held in **organised form** (files, paper or electronic). The Act does not cover invasions of privacy that don’t use/misuse *information* (someone peering through your window); does not cover privacy invasions that do not use *organised* information (twitter); it also does not cover invasions of privacy that use organised information held for *domestic* use (your address book) [3]. The DPA ostensibly provides a partial but focused approach to privacy protection, intended to prevent the sorts of invasions of privacy that work by searching or mining files and databases. Personal information is only to be used where the data subject consents, and use for further purposes requires new consent.

2. Data Protection Regulates Content not Acts

DPA is unusual legislation. It aims to regulate all action that uses *specific types of information or content*, rather than regulating specific *types of action*. It covers what it terms the ‘processing’ of *any* information that is personal, or both *personal* and

[1] Art. 8 protects "private and family life, his home and his correspondence”

[2] European Directive 95/46/EC. The directive states that ‘‘personal data' shall mean any information relating to an identified or identifiable natural person ('data subject')’ (Chapter 1, Article 2 (a)).

[3] In a world in which many address books held and compiled by individuals contain linkable lists of contacts with extensive biodata it is not clear to me whether a convincing distinction between an address book and a database can be drawn.

sensitive, [4] and prohibits such processing without consent. The Act therefore requires a clear distinction between personal and non-personal data. It defines personal data as

> …data which relate to a living individual who can be identified (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller.[5]

It then links this account of personal data to a wide definition of *processing* that covers acquiring, recording, organizing, altering, retrieving, linking, consulting or using data by whatever means. The *Legal Guidance* to the Act states that 'The definition [of 'processing'] in the Act is a compendious definition and it is difficult to envisage any action involving data which does not amount to processing within this definition'. [6]

Three questions arise. What are 'data that relate to a living individual'? When do data and other information make an individual identifiable? What sort of consent to a reuse of personal data is needed for it to be lawful?

3. "Relating to a Living Individual who can be identified"

The term 'data relating to a living individual' is magnificently obscure. It does not mean simply data that are true of (living) persons. Much that is true of each of us is general, not personal information, and true of many or all others. Each of us has human ancestors and was born at some time in the past. DPA does not regulate this information. On the other hand, personal information does not have to be uniquely true of persons. Many diseases are common, yet the health records of those who suffer them are personal.

The crucial element in the Act's definition of personal information is the idea of information that makes an individual *identifiable* on the basis of 'other information' held by, or likely to be held by the data controller, or by others. Since that other information will vary, the *inferences* people can draw from (supposedly) personal information vary, as will the identifications they can make. A small piece of information may make a person identifiable to those with the right other information, as every reader of detective stories knows.

[4] Data are *sensitive* as well as *personal* if, for example, they include data about racial or ethnic origin, physical or mental health or sexual life. See DPA 1998, (2) and DoH Guidance on records management 2009 http://www.dh.gov.uk/en/Managingyourorganisation/Informationpolicy/Recordsmanagement/DH_4000489 .

[5] *Data Protection Act 1998*, Part I, Section 1. This formulation is closely based on that of the European Directive 95/46/EC which states that ''personal data' shall mean any information relating to an identified or identifiable natural person ('data subject')' (Chapter 1, Article 2 (a)).

[6] See *Data Protection Act 1998: Legal Guidance.*, p. 15 Data Protection Act 1998 Legal Guidance.pdf

4. Identifiability and Anonymisation

Attempts to nail down out what makes individuals 'identifiable' have led to controversy, particularly in medical and social research. On one view, if data are *anonymised* they are non-personal. However where anonymisation is reversible, however secure the encryption and however limited access to the key, individuals will be *in principle* identifiable by *some* means by *some* persons. This leads some to conclude that reversible anonymisation cannot satisfy data protection requirements. The situation is only worsened by data mining techniques that may enable identification of individuals even when data are (ostensibly) irreversibly anonymised. There is no general agreement on standards for anonymisation for reuse of personal data, and this obstructs and burdens all secondary data analysis and public health work. Some think even the highest standards of encryption are not enough; others think it odd to demand higher standards of data protection than those achieved in the use of non organised information in medical settings.

5. Consent and Reconsenting: Commerce vs. Biomedicine.

Lawful reuse of personal data, it seems requires consent from data subjects, or a case for exceptional treatment. Consent requirements are understood in radically different ways in commercial and in biomedical contexts. In commercial life we consent to the reuse of personal data for sundry purposes almost without noticing it. We sign forms that we do not read, we click and tick and 'accept' complex contracts and terms and conditions. It would be great exaggeration to pretend that this amounts to *informed* consent: but it is taken to make specified further uses of personal data lawful.

These convenient fictions of consent are not available in biomedical contexts, where standards for informed consent are more exacting, sometimes impossibly exacting. It is impossible to seek consent to unforeseen future research uses at the point of treatment, and often impractical to seek new consent when research projects are formulated, or new analysis undertaken. This burdens medical and social research: all work in public health and epidemiology, including all secondary data analysis, seemingly needs specific consent to further use from each data subject—although the data are to be reused for *impersonal* ends. Obtaining further consent from all source subjects is often impractical; selective reconsenting is likely to damage research findings by skewing their statistical basis.

6. Making Exceptions

If consent cannot be obtained, one alternative would be making a case for exceptional treatment. Some forms of unconsented to reuse are permitted by DPA, for example reuse for clinical audit. A system for granting special permissions was set up in the *Health and Social Care Act 2001)* administered at first by the *Patient Information Advisory Group* (PIAG), now by the *National Information Governance Board* (NIGB).

The system has been subject to critical reports by the Academy of Medical Sciences in 2006,[7] as well as by the Information Commissioner together with the Director of the Wellcome Trust in 2008, who concluded that:

[7] Academy of Medical Sciences, *Personal data for public good: using health information in medical research*, 2006

> It is clear that the framework as it stands is deeply confusing and that many practitioners who make decisions on a daily basis about whether or not to share personal information do so in a climate of considerable uncertainty. [8]

They recommended as an alternative that:

> 'Safe havens' should be developed as an environment for population-based research and statistical analysis in which the risk of identifying individuals is minimised; and that a system of approving or accrediting researchers who meet the relevant criteria to work within those safe havens is established. [9]

The NHS avoids some difficulties by being the data controller for all patient data, so eliminating the appearance of data sharing. It can even incorporate researchers who are not employees by making them honorary consultants. But parallel moves are not available for social research, and not always convincing for medical research. It is I think an open question whether any system of exemptions for evading the effects of DPA would be ethically acceptable or effective way to control access to reversibly anonymised patient information for research purposes.

7. Defective Legislation

I want now to cut to the chase and state why I have think that piecemeal remedies that retain the framework of the current DPA are likely to fail. I begin by noting that we live in a world where each patient is treated on the basis of information gained by treating others, and would be horrified if this were not so. Any account of privacy in biomedical contexts must recognize this reality: information sharing is not some foul disease to be avoided. It is basic to medical practice, and obstructing it, even with the best of intentions, is unacceptable.

The basic problem of DPA 1998 is that it assumes that data can be partitioned into the personal and the non personal. This is false. If the receptionist at your GP calls out your home address she discloses personal (not sensitive) information, which the surgery hold for purposes connected with your medical care, but should not communicate to others or allow others to overhear without your prior consent. Yet on the other side of town there is an electoral register containing your name and address, for the public and for political parties to consult without seeking your consent. So are your name and home address personal information? Are they subject to Data Protection? Or does it depend on the context? If it depends on context, then the basic assumption that data can be divided into the personal and the non personal fails.

Because its fundamental assumption that data can be partitioned into the personal and the non personal fails, the Act is breached everyday. Mostly and sensibly we overlook these breaches. Consider these examples:

1. A doctor *revisits* information about the treatment of past patients in order to refresh her knowledge before treating a current patient without explicit consent from each past patient. Acceptable or not?
2. A doctor takes a family history, recording presumed personal information about a relative, without prior consent? Acceptable or not?
3. A doctor writes a case note, using information originally provided for treatment for other purposes? Acceptable or not?

[8] Richard Thomas and Mark Walport, *Data Sharing Review Report,* 2009 para 5.
[9] Recommendation 15

Each of these everyday stories is about the disclosure or use of information for purposes other than those for which it was first obtained in a medical setting. Do we really want these activities to be unlawful without renewed consent? Or do we need to find a better basis for protecting informational privacy?

8. Would Confidentiality Serve us Better?

Before the days of data protection, confidentiality was taken as a fundamental norm for using information provided in medical contexts. Confidentiality governs *types of action* — specifically *types of speech act* — and does not aim to regulate all 'processing' of *types of information or content*.

An approach to informational privacy based on extending the law of confidentiality would not require anyone to determine which information is or is not *personal*, or *personal and sensitive*, or what it takes to make individuals identifiable to others. Rather than demanding that we first define and then protect all 'personal' content, confidentiality offers a way of protecting content *of any type* that the parties to a communicative transaction seek to protect, have agreed to protect, or are required to protect. It can be invoked for specific aspects of professional, commercial or other relationships, and can once again be waived by seeking consent from the confider. The central difference is that in imparting confidential information the recipient takes on obligations not to share the information without sharing the relevant obligation. There is much more to be said here.

9. Some Conclusions

Data protection legislation has created substantial difficulties for medical and social research, without providing good protection for informational privacy. It creates particular difficulties for impersonal secondary use of legitimately acquired, lawfully held data by requiring specific reconsenting.

Given these realities, it is of little help that data protection legislation permits reuse of data when all subjects consent, or exceptionally if permission can be obtained from a statutory body. Obtaining further consent from all subjects is often impractical; selective reconsenting is likely to damage research findings by skewing their statistical basis; seeking exemptions is demanding, time consuming, and according to anecdote erratically successful.

I conclude that there are good reasons for revised data protection legislation. The redrafting of the European Directive, on which consultations are proceeding, and any subsequent legislative changes in the UK, need to be quite radical if informational privacy is to be served. I believe that in seeking reform it would be better to focus on regulating the acts by with content is communicated and not the 'processing' of ill-defined types of informational content.